# Bayes Academy - An Educational Game for Learning Bayesian Networks

Kaj Sotala

Master's thesis
UNIVERSITY OF HELSINKI
Department of Computer Science

Helsinki, October 20, 2015

HELSINGIN YLIOPISTO — HELSINGFORS UNIVERSITET — UNIVERSITY OF HELSINKI

| Tiedekunta — Fakultet — Faculty | Laitos — Institution — Department |
|---|---|
| Faculty of Science | Department of Computer Science |

Tekijä — Författare — Author
Kaj Sotala

Työn nimi — Arbetets titel — Title
Bayes Academy - An Educational Game for Learning Bayesian Networks

Oppiaine — Läroämne — Subject
Computer Science

| Työn laji — Arbetets art — Level | Aika — Datum — Month and year | Sivumäärä — Sidoantal — Number of pages |
|---|---|---|
| Master's thesis | October 20, 2015 | 64 |


Tiivistelmä — Referat — Abstract

This thesis describes the development of "Bayes Academy", an educational game which aims to teach an understanding of Bayesian networks.

A Bayesian network is a directed acyclic graph describing a joint probability distribution function over $n$ random variables, where each node in the graph represents a random variable. To find a way to turn this subject into an interesting game, this work draws on the theoretical background of meaningful play. Among other requirements, actions in the game need to affect the game experience not only on the immediate moment, but also during later points in the game. This is accomplished by structuring the game as a series of minigames where observing the value of a variable consumes "energy points", a resource whose use the player needs to optimize as the pool of points is shared across individual minigames. The goal of the game is to maximize the amount of "experience points" earned by minimizing the uncertainty in the networks that are presented to the player, which in turn requires a basic understanding of Bayesian networks.

The game was empirically tested on online volunteers who were asked to fill a survey measuring their understanding of Bayesian networks both before and after playing the game. Players demonstrated an increased understanding of Bayesian networks after playing the game, in a manner that suggested a successful transfer of learning from the game to a more general context. The learning benefits were gained despite the players generally not finding the game particularly fun.

ACM Computing Classification System (CCS):

- **Applied computing - Computer games**
- **Applied computing - Interactive learning environments**
- *Mathematics of computing - Bayesian networks*


Avainsanat — Nyckelord — Keywords
bayesian networks, computer games, computer science education, educational games, serious games

Säilytyspaikka — Förvaringsställe — Where deposited

Muita tietoja — Övriga uppgifter — Additional information

# Contents

# List of Figures

# 1 Introduction

There is currently a strong interest in using games for educational purposes, both in academia (Shaffer, 2006; Gee, 2007; Whitton, 2010) and among the general public. In 2013, the Finnish Minister of Education announced her intention to gather a group of game designers to provide advice on how to develop mathematics education (MTV3, 2013). Educational games were also highlighted in a report issued by the Finnish Parliament's Committee for the Future (Eduskunnan tulevaisuusvaliokunta, 2013). In the United States, President Barack Obama has also expressed his desire to see improved educational games (Farr, 2013).

Educational games are not only being talked about, they are also being played. In June of 2013, 4000 K-12 students in Washington State engaged in an ”Algebra Challenge”, solving close to 400,000 equations by playing a game (Washington State, 2013), and in July, the game-like language learning service Duolingo claimed to have more than five million active users (Lardinois, 2013).

This thesis sets out to develop an educational game, “Bayes Academy”, which aims to teach the general ideas of *Bayesian reasoning* as well as the specific mathematical framework of *Bayesian networks.*

## 1.1 Serious games

Games have a long history of being used for education, particularly in the military, where sophisticated wargames were developed for training officers (Michael & Chen, 2005). Business games (Greco et al., 2013), used for teaching business management, have been very widely used by both companies and business schools since the 1950s (Faria, 1998).

One reason for using games for education is that even pure entertainment games require considerable learning before they can be mastered (Gee, 2007; Koster, 2010). Documented effects from first-person shooters include an improved ability to multitask, differentiate targets, prioritize different targets, and work in a team with minimal communication (Michael & Chen, 2005, p. 58-59). Experienced players of online poker seem to have better emotional control and be capable of better coping with losses inflicted by bad luck (Palomäki, 2013). In one case, a computer game may even have taught knowledge which saved a person's life: James Sterrett reports that his brother recognized the symptoms of appendicitis due to having played the surgery game Life and Death as a child, which caused him to seek medical care instead of ignoring the symptoms (Sterrett, 2012).

Games designed for entertainment have also been used for explicit educational purposes. Squire (2004) ran three studies in which he used a strategy game to teach students history. In the first study, the students were made up of a class where everyone had failed at least one course, and who were originally very disinterested and hostile to any teaching. Even though they found the game too complex and challenging at first, they gradually came to enjoy it and started drawing connections between the gameplay and real history. Over time, their questions and analysis of the game became increasingly sophisticated, and they began balancing multiple variables as well as critically evaluating the game's historical realism.

If games can have powerful learning benefits even when they are not designed for the purpose of education, it stands to reason that explicitly educational games could do even better. Serious games are

> ...games that do not have entertainment, enjoyment, or fun as their primary purpose. That isn't to say that the games under the serious games umbrella *aren't* entertaining, enjoyable, or fun. It's just that there is another purpose, an ulterior motive in a

very real sense. (Michael & Chen, 2005, p. 21)

Besides education, serious games can also be designed for purposes such as exercise, therapy, politics, or marketing. As an example of a marketing game, *America's Army* (U.S. Army & Secret Level, 2002) places the players in a realistic simulation of serving in the US Army. One Army study found that America's Army had given people in the 16-24 year old age bracket a stronger positive impression than any other recruitment technique (Michael & Chen, 2005, p. 55).

A related term is "gamification", defined by Deterding et al. (2011) as

> the use of game design elements in non-game contexts.

Marczewski (2013) distinguishes between gameful design, gamification, serious game / simulation, and game. In this classification, the category of "gameful design" is said to include services such as Twitter, which include an element of playfulness, such as by showing a "fail whale" in their error messages. Marczewski defines gamification as

> what you get when you take elements and ideas from games and apply them to things that are not games. So adding progress bars to a site to show how much of your profile you have filled in (e.g. linkedin.com), adding points, badges, leaderboards, peer pressure and more to things that normally would not have them (e.g. Nike+, Idea Nation, Zombies Run, Gamification.co).

Serious games are products that "look and feel like a real game" but are developed for some purpose other than entertainment, whereas the category of games includes the products that contain all of the above elements, but are made for mere fun. The focus on this thesis is on what Marczewski's classifcation would refer to as "serious games".[1]

[1] Several authors have offered different definitions for the word "game", including but not limited to Huizinga (1955), Caillois (1961), Salen & Zimmerman (2004) and Schell (2008).

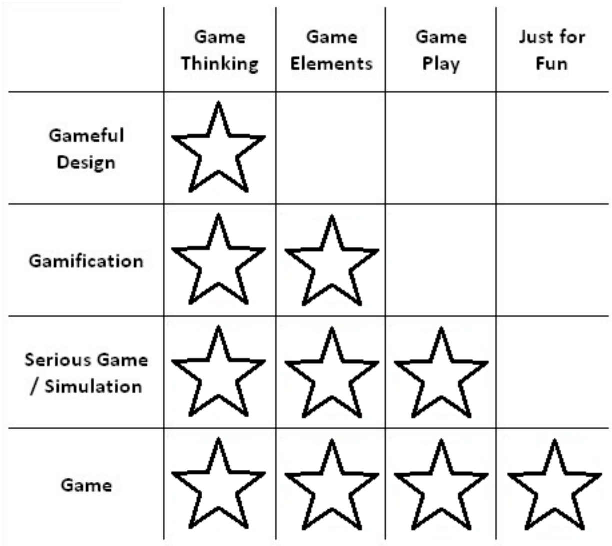



Figure 1: *Types of design: gameful design, gamification, serious game, game (Marczewski, 2013).*

# 2 Bayesian reasoning

*Bayesian reasoning* or *Bayesian epistemology* (Pearl, 1988; Hájek, 2012; Talbott, 2013) refers to a philosophy in which probabilities are understood as subjective degrees of belief, and where there exist formal mathematical rules for modifying one's beliefs in response to new evidence using Bayes' Theorem:

$$P(X|Y) = \frac{P(Y|X)P(X)}{P(Y)}$$

Bayesian reasoning and techniques have been deployed in a variety of fields, such as criminology (Boondao, 2008), cognitive science (Chater et al., 2006; Jones & Love, 2011), ecological modeling (Uusitalo, 2007), education (Sucar & Noguez, 2008), forensic science (Garbolino & Taroni, 2002), genetics

This thesis chooses to avoid the question altogether, using the word in its common-sense meaning without trying to offer an exact definition.

(Sebastiani & Perls, 2008), information retrieval (Campos et al., 2008), medicine (Nicholson et al., 2008; Onisko, 2008), modeling of mineral potential (Porwal & Carranza, 2008), neuroscience (Kording, 2007), philosophy of science (Howson & Urbach, 1993), and wine classification (Duarte-Mermoud et al., 2008).

## 2.1 Bayesian networks

A specific technique associated with Bayesian reasoning, though not necessarily requiring commitment to a Bayesian philosophy, is that of Bayesian networks (Pearl, 1988). A Bayesian network is a way of representing the conditional and independence relationships between a group of random variables: informally, it is a way for reasoning about some thing X which we are interested in but might not be able to observe directly, but which is associated with another thing Y which we can observe.

For example, a network involving the probability of a burglary, an earthquake, and a burglar alarm going off may let us determine whether the alarm going off (something we observe) is indicative of a) our home having been robbed or of b) there having been an earthquake (possible causes for the alarm that we cannot immediately observe if we are not present).

Formally, a Bayesian network is a directed acyclic graph describing a joint probability distribution function over $n$ random variables. Each vertex $X$ in the graph represents a random variable that can take two or more possible values. Edges between the vertices represent influences (either causal or not, depending on the interpretation) between the variables. Given the set $\Pi_{X_i}$ of vertices that are the parents of $X_i$, there is a table $P(x_i|\Pi_{X_i})$ which gives the conditional probabilities of the event $X_i = x_i$, given any value combination $\Pi_{X_i}$ of the parent set $\Pi_{X_i}$. Combining these tables gives an overall joint distribution function

$$P(x_1, x_2, ..., x_n) = \prod_{i=1}^{n} P(x_i|\Pi_{X_i})$$

In addition to its academic usage, Bayesian reasoning is also highly relevant in people's daily lives. For example, various argumentative "fallacies" can be understood as arguments that are flawed because in Bayesian terms, they only provide weak rather than strong evidence (Hahn & Oaksford, 2007). Bayesian theory can also help explain the function of various everyday events, such as the experience of being surprised by something (Itti & Baldi, 2005).

Because of its general and wide-reaching applicability, Bayesian reasoning seems particularly suited for being taught via a game, as it can be applied in a context which the learner already finds inherently compelling. To teach Bayesian reasoning through a game, the designer does not need to create a game about an academic topic that might seem remote from most people's lives. Rather the designer can take a domain of which there already exist entertainment games, and look for ways to bring out the Bayesian elements involved in acting in that domain.

## 2.2 Teaching Bayesian reasoning

Previous work on teaching Bayesian reasoning has focused on ways to teach the application of Bayes' theorem in what corresponds to a simple two-node Bayesian network. A typical example question[2] gives the following facts:

- The probability that a woman who undergoes mammography will have breast cancer is 1%.
- If a woman undergoing mammography has breast cancer, the probability that she will test positive is 80%.

[2] This example is inspired by Eddy (1982), who found that most doctors incorrectly gave answers around 70-80%, when the real answer was approximately 8%.

- If a woman undergoing mammography does not have cancer, the probability that she will test positive is 10%.

The subjects are then asked, given that a woman tests positive, what is the probability that she has cancer?

The most important results of the studies have involved the effect of natural frequencies and visualizations on teaching Bayesian reasoning.

### 2.2.1 Natural frequencies

Research has found that presenting probabilities in terms of "natural frequencies" facilitates the understanding of Bayesian reasoning (Gigerenzer & Hoffrage, 1995; Sedlmeier & Gigerenzer, 2001; Hoffrage et al., 2002). A natural frequency is defined (Gigerenzer & Hoffrage, 1995; Hoffrage et al., 2002) as one that

1. has been obtained by natural sampling, such as by observing a real population and noting the prevalence of a disease and its symptoms in that population.

   An example of such an estimate is "there are 1,000 people, 10 of whom have the disease. 8 people with the disease have the symptom, and out of the 990 without the disease, 95 also have the symptom".

   In contrast, a number obtained from a population where the base rates have been artificially fixed, such as "out of 1,000 people with the disease, 800 have the symptom, and out of 1,000 people without the disease, 96 do not", is not the result of natural sampling.

2. has not been normalized into a percentage, such as by dividing 95 by 990 to get 0.096.

If natural frequencies are used, the probability of a person being sick given that they have the symptom can be calculated simply as

$$P(Sick|Symptom) =$$
$$\frac{Sick \wedge Symptom}{(Sick \wedge Symptom + Healthy \wedge Symptom)} = \frac{8}{8+95} = 0.078$$

However, if natural frequencies are not used and the information is expressed as “the probability of a sick person having the symptom is .8 and the probability of a healthy person having the symptom is .096”, then the information about the ratio of sick people with the symptoms to all people with the symptoms is lost. It will then need to be recalculated using the formula

$$P(Sick|Symptom) =$$
$$\frac{P(Sick)P(Symptom|Sick)}{P(Sick)P(Symptom|Sick) + P(\neg Sick)P(Symptom|\neg Sick)}$$

which will arrive at the same probability, but requires considerably more calculation and is harder to understand and remember.

### 2.2.2 Visualizations

When people are asked to solve problems involving probabilities, they often use various visual representations of the problem as aids (Zahner & Corter, 2010). Providing people with the right visual aids also helps them solve problems better (ibid). This suggests that the right visualizations could be an important tool for helping people learn probabilities.

Several studies have considered the impact of visualization on learning probabilities. Sedlmeier & Gigerenzer (2001) found a slight advantage to presentation where the structure of the problem was broken down to a tree, but did not consider it significant. Sedlmeier (2000), in teaching conditional probabilities, found that a grid representation provided a substantial benefit over a Venn diagram representation. Sedlmeier (1997) found representational training with probability trees and grid representations to substantially outperform rule training, but did not find a major difference between grid and Venn representations. Micallef et al. (2012) tested a number of different visualizations on a population recruited from Amazon Mechanical Turk and found that visualizations provided no clear problem-solving benefit when they were simply added as extras on top of a Bayesian verbal problem. However, providing the statistical information relevant for the problem in the visualizations alone, provided better results than either providing the numbers in text *or* than providing the same numbers in both text and pictures.

# 3 Theoretical background for game design

The general design of the game draws on two theoretical backgrounds: meaningful play (Sec. 3.1) and cognitive apprenticeship (Sec. 3.2).

## 3.1 Meaningful play

Educational games have often suffered from the problem of being "chocolate-coated broccoli" (Bruckman, 1999), where players are provided with brief rewards in exchange for playing through long, mostly uninteresting educational sections. In contrast, games made with pure entertainment in mind may be very challenging, requiring the player to develop very high levels of skill (Gee, 2007). Mastering the game may take intense effort and repeated attempts, during which the player might never feel bored, for the activities in the game remain meaningful throughout.

Salen & Zimmerman (2004) define the goal of successful game design to be the creation of psychologically and emotionally *meaningful play*, which is said to occur when the actions in the game are both *discernable* and *integrated* into the larger context of the game.

By *discernable*, the authors mean that the results of a player's actions, as well as any other pieces of information that are relevant to the player, are communicated in a perceivable way. When the player takes an action, they must know what happened, rather than experience the game as one of random buttom-mashing.

An action being *integrated* to the larger context of the game means that

> an action a player takes not only has immediate significance in the game, but also affects the play experience at a later point in the game. Chess is a deep and meaningful game because the delicate opening moves directly result in the complex trajectories of the middle game—and the middle game grows into the spare and powerful encounters of the end game. Any action taken at one moment will affect possible actions at later moments.

When an entertainment game is teaching its players how to play, the learning content is experienced as meaningful rather than uninteresting because it helps the player to better take actions that are meaningful in this sense. Rather than having rewards with no logical connection to the educational content, a good entertainment game allows a player to use what they have learned to make better decisions, which then influence the overall game.

For example, in the turn-based tactical shooter XCOM: Enemy Unknown by Firaxis Games (2013), the player is faced with choices such as whether to shoot or throw a grenade at an enemy that has three remaining "hit points". If the player throws the grenade, the enemy is certain to die, but the player loses the grenade and can no longer use it for the rest of the mission. If the player rather chooses to shoot at the enemy, they may keep the grenade and use it later, but there is a chance of missing, allowing the enemy to shoot back at the next turn.

Whether or not this is an acceptable risk depends on the probability of the player's soldier hitting, the damage that their weapon is capable of dealing as well as the expected damage on an average hit, the number of remaining aliens on the mission, the amount of grenades that other soldiers in the player's squad have, and so on. Solving this dilemma involves a number of arithmetic and probability calculations, whose results can then be used to estimate the likely outcome of the different decisions that are possible in the situation. The immediate consequences of the choice will influence the outcome of the rest of the mission, and the outcome of the rest of the mission will in turn influence the player's options are performance in the rest of the game.

This suggests a way to make the game developed in this thesis interesting and enjoyable. As in entertainment games, learning the content must lead to clear improvements in the player's ability to evaluate their options and to make good choices. The outcomes of these choices, in turn, must have a clear and discernable impact on the way that the events unfold in the remaining game.

## 3.2 Cognitive apprenticeship

*Cognitive apprenticeship* (Collins, 1988; Collins et al., 1991; Collins, 2006) is an approach to instruction which seeks to "make thinking visible". In traditional apprenticeship, a master would show their apprentice how to perform a task, watch the apprentice perform the same task and provide feedback, and then gradually turn over more and more tasks for the apprentice as their skills matured.

When domains such as reading, writing, and mathematics are being taught, the process is typically not clearly visible the way that it is in traditional apprenticeship. The teacher cannot see the thinking process underlying the work of the students, and the students cannot see how the teacher produces their work. Cognitive apprenticeship seeks to externalize the thought processes in such a way that all involved parties become capable of perceiving it, thus learning from the processes of others and becoming capable of providing feedback to others. For example, students learning reading strategies may be asked to formulate questions about the text, summarize it, clarify it, and make predictions of what will come next. In this way, students gradually learn to do more than just passively read the text, while also needing less and less coaching ("scaffolding") from the teacher. Explicitly decomposing a broad, abstract task such as reading into concrete subtasks helps the learner to better understand the steps needed to solve the problem.

An educational game offers several opportunities to implement these principles. In the early stages of the game, the player may be provided with clearly labelled and illustrated demonstrations of the learning material, with the player initially only needing to carry out a relatively small number of simple tasks. When the player learns the basic principles, the amount of scaffolding is gradually reduced. Eventually they may progress to carrying out very sophisticated tasks themselves.

# 4 Design goals

Designing an educational game requires great care. On the other hand, educational games have a reputation of not being very fun (Bruckman, 1999). On the other, games may also suffer from a stigma of being viewed as uneducational. In one attempt at using dice games for teaching probability, many students did not take the games seriously and asked questions such as "When are we going to do real maths?" and "Are we still playing games?" (Bennie, 1998). If the designer is not careful, a game may easily come across as juvenile, biased, boring, or unrealistic (Michael & Chen, 2005, p. 211).

## 4.1 Educational goals

The design of an educational game should start out with the instructional goals (Michael & Chen, 2005, p. 95).

Knowledge of some subject consists of both conceptual and procedural knowledge. Rittle-Johnson et al. (2001) define procedural knowledge as "the ability to execute action sequences to solve problems", noting that such knowledge is often "tied to specific problem types and therefore not widely generalizable". In contrast, conceptual knowledge is defined as "implict or explicit understanding of the principles that govern a domain and of the interrelations between units of knowledge in a domain". They argue that these two types of knowledge develop bidirectionally, with a learner's improved procedural knowledge improving their ability to learn more conceptual knowledge, and vice versa.

This work began in part from the author's fascination with a specific kind of conceptual knowledge relating to Bayesian reasoning. It was his experience that elements of Bayesian reasoning are ubiquitous in daily life, and that these underlying principles could be seen nearly everywhere. While comprehensively teaching the conceptual knowledge is outside the scope of this work, having the player learn some of the basic procedural skills underlying the conceptual knowledge seem like a useful goal. The main skill that this work focuses on teaching is a basic understanding of the flow of probability in a Bayesian network, operationalized as a test of understanding d-separation (Koller & Friedman, 2009, p. 69).

Schell (2008) argues that the purpose of games is to deliver experiences, and that part of the game design process is to decide on the kind of experience that the designer wants to deliver. While Schell (2008) is writing in the context of designing video games for entertainment, Squire (2006) takes an education-focused perspective and similarly suggests viewing games as designed experiences, "in which participants learn through a grammar of *doing* and *being*". The ideal experience to convey to the player would be one where the player came to see the world through a Bayesian lens.

## 4.2 Target audience

Consideration of the target audience is an important component in game design (Schell, 2008).

Not all players have the same preferences, and different players will like different games. Heeter et al. (2011) recruited 330 undergraduates interested in obtaining extra credit in telecommunication or history classes at a large Midwestern university, and had them play four different games. Even in this relatively homogenous group, no game was universally liked nor universally disliked. Instead, each game had both eager players and resistant players.

While it is often assumed that most people find games intrinsically motivating, the motivations for playing games vary, and many people consider themselves non-gamers. Whitton (2007) found that people who described themselves as gamers attributed their motivation to either intellectual challenge ("Cerebral Gamers"), social interaction ("Social Gamers"), or exercise and physical exertion ("Physical Gamers"). Gamers often prefer games that appeal to their strengths (Sherry & Dibble, 2010), and cautious players prefer different levels of risk than adventorous players (Graesser et al., 2010).

People who described themselves as non-gamers would also sometimes play games, but tended to only do so in two situations: to use the games as a social facilitator, such as a way to getting to know people, or for brief periods when bored (Whitton, 2007). Whitton also found that even in a group of predominantely male, young computing students, fewer than two thirds of the students felt that games would by themselves be a particularly motivating way to learn.

Non-gamers, who might not be as skilled with games as expert gamers are, may need to be taken into account by making it as easy possible to start playing the game. In one study, an undergraduate university student who was assigned to play a game for ten minutes quit before the time was up, saying that "I did not know what to do. I hate playing video/computer games." (Heeter et al., 2011) Similarly, Whitton (2007) identified four factors that were felt to be demotivating by gamers and non-gamers alike: difficulties in getting started, getting stuck in the activity, lack of trust in the environment, and intrinsic boredom with the subject matter or game itself.

For the game developed in this thesis, the target audience is that of people with a relatively technical mindset who might already have an interest in probabilistic reasoning, but no previous experience with Bayesian networks. While aiming for a broader target audience would be valuable, it would also be considerably more challenging.

## 4.3 Fun

The question of whether educational games should be fun is somewhat contentious (Michael & Chen, 2005, p. 40). Deep learning of a subject may reduce enjoyment, as it requires considerable effort. Being confused forces a player to think and is thus linked to learning, but being confused is not necessarily fun and may break immersion. (Graesser et al., 2010) An educational game which attempted to provide a maximally fun experience might be forced to sacrifice some educational value.

On the other hand, many entertainment games are also very challenging, and require substantial effort in order to be enjoyed. Michael & Chen (2005) point out the popularity of documentaries and non-fiction, and argue that people are inherently interested in things that they can relate to and which tell them interesting things about their lives. Thus, if people feel that a game teaches them valuable things, they may be motivated to play it even if it was not terribly fun.

Regardless, I hold that the game should aim to be fun. Unless people are forced to play the game in as setting such as school, it needs to rely on its intrinsic appeal if it is to be played at all. The game should be sufficiently entertaining that people are willing to play it for the entertainment value, even if they were not interested in the learning content for its sake.

Another reason to make the game fun relates to the question of transfer of learning (Macaulay, 2001), and whether practice at a particular task carries over to real life. Transfer has often been considered difficult to achieve, with researchers documenting various cases of people being good at solving mathematical problems either in school or in their regular life, but not vice versa (Lave, 1988).

Various video games may provide for transfer without even intentionally attempting to do so. People who spend considerable amounts of time playing visual video games see images of the game as they are falling asleep (Stickgold, 2000), and there are many anecdotal reports of people playing a game and then automatically thinking in terms of that game in real-life situations (for one list of examples, see (TVTropes, 2014)). One could thus conjecture that a game that was played sufficiently could achieve transfer by the simple amount of repetition, assuming that its content was thus that could feasibly be transferred to real-life situations.

# 5 Bayes Academy

Many math games, such as Algeburst (Cengage Learning & Muzzy Lane Software, 2012) and DragonBox (WeWantToKnow, 2012) are essentially pure puzzle games in which the player solves math problems in order to proceed in the game, but the tasks do not try to explicitly model any real-world situation. My initial plan was to instead situate all the puzzles into a narrative in which all the levels would model the player character doing something in the world. This would have given the player a better understanding of how the mathematics of Bayesian reasoning are applicable in real life. However, due to time constraints I eventually discarded most of the plotline, and went for what was essentially a pure puzzle game instead.

The game, Bayes Academy, is essentially a resource management challenge. The player is shown various Bayes nets, and needs to reduce the uncertainty in the network as much as possible by discovering the true values of the different nodes. This is done by clicking on the various nodes, which observes their true value. However, each observation costs a point of "energy", an in-game resource of which there is only a small amount available. The player needs to find the nodes which produce the most information when observed, to get the most out of the expended energy. If the player is not sufficiently effective in using their energy, they will lose the game.

The game is implemented in Java and its source code is available online[3] under the Apache license.

## 5.1 Bayes networks in the game

In the game's representation, most of the nodes in the game are either "prior nodes", that have some fixed probability of being true or false, or "conditional

[3]https://github.com/ksotala/BayesGame/

probability nodes" whose value is deterministically derived from the values of their parent nodes. Conditional probability nodes include basic logical operations such as AND, OR, and IS, with no randomness involved. This choice was made in order to make it easier for the player to learn the system. By making their values deterministic, the player can notice that, for example, AND is always true when both of its parents are true, and false otherwise.

Even though the values of the conditional probability nodes are determined deterministically, the player's uncertainty about the values of the node parents still allows probabilistic reasoning to be used. For example, if both parent nodes for an AND node are prior nodes that have an independent .5 chance of being true, then the there is a $(0.5 \cdot 0.5 =)$ 0.25 chance for the AND node to be true. The game displays these probabilities graphically. All of the levels are implemented as genuine Bayesian networks, and as the player makes observations, the networks are updated with an implementation of the sum-product message-passing algorithm (Koller & Friedman, 2009, p. 357).

A more challenging type of nodes that appear in the game are the stochastic "Bayes nodes", which have a certain probability of being true if the parent node is true, and some other probability of being true if the parent node is false.

### 5.1.1 Prior and IS nodes, observability

In an early part of the game's introductory tutorial, there is a very simple "level" (left side of Fig. 2), in which the player is presented with a two-node Bayesian network. The top node is a prior node, defined as a node that has no parents and which is true with some probability that is independent of any other nodes. This is symbolized by the two squares at the top, showing that it can be true or false independently of anything else. Its probability, in this case 50%, is shown graphically in the colored grid below the two

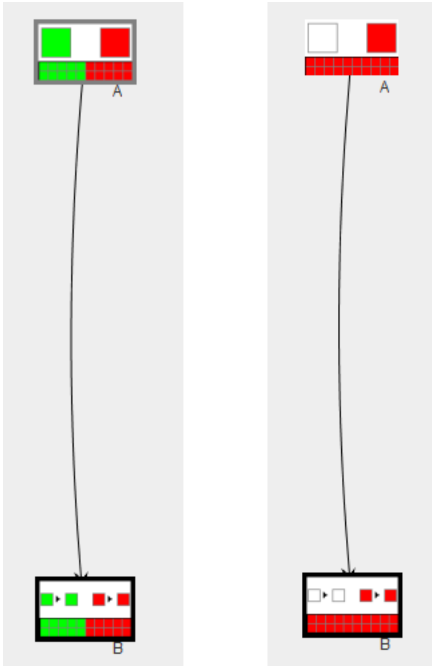



Figure 2: *An early tutorial level, consisting of a prior node (top) and an IS node (bottom). Both are initially shown as having a 50-50 probability (left); after the player clicks on the "A" node to inspect it, both the prior node and the IS node are shown as being false (right).*

squares[4].

The bottom square is an IS node, which simply takes the same value as its parent node. This is symbolized by the two pairs of boxes at the top of the IS node, representing a parent node and a child node which have the same truth values.

On the first tutorial level, the prior node is surrounded by a grey border and the IS node is surrounded by a black one. This signifies that the prior node is *observable* and the IS node is an *unobservable*: the player can click on the prior node to observe it and thus find out its truth value, but clicking on

[4]The game's default color scheme codes truth as green and non-truth as red. However, alternate color schemes for color-blind users are also provided.

the IS node does nothing. However, when the player does click on the prior node, the resulting graph (right side of Fig. 2) also updates the value of the IS node based on the value of the prior node. Once an observable node is observed, it loses its grey border.

#### 5.1.2 AND nodes and OR nodes

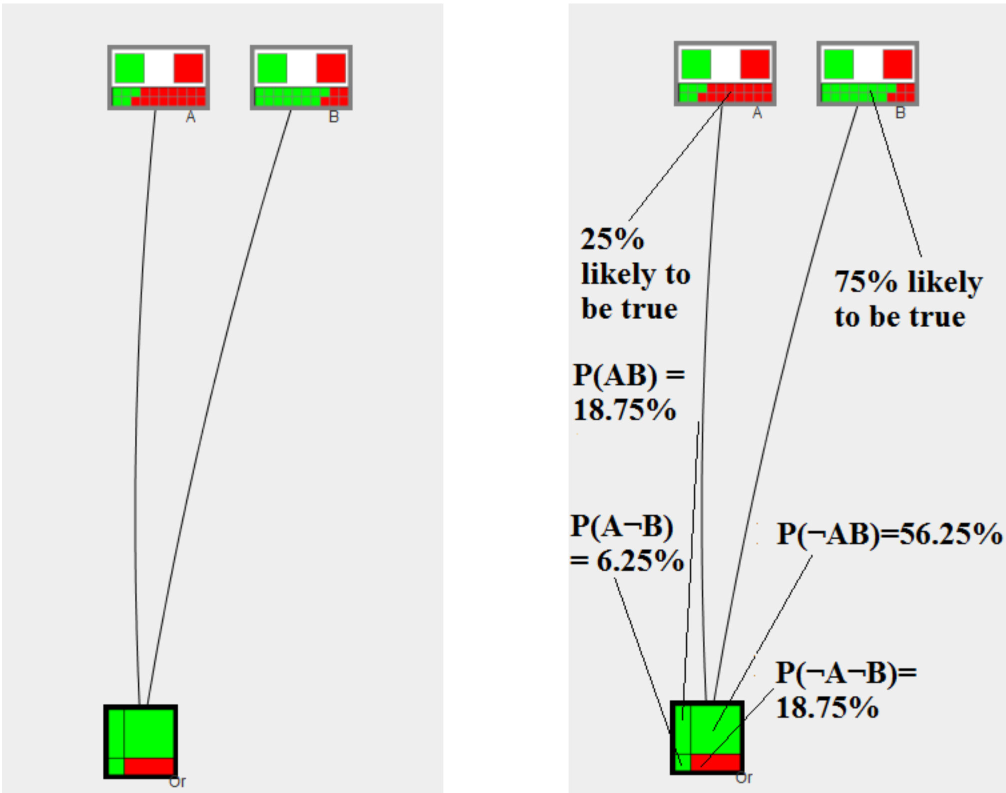



Figure 3: *An OR node with two parents A and B, with respective 25% and 75% probabilities. The rectangles inside the OR node represent the various combinations of values that its parents can take, and the surface areas of the squares reflect the relative probabilities of those combinations. Squares are colored red or green depending on whether they would make the overall node evaluate as true or false.*

OR and AND nodes always have exactly two parents, and a 2x2 grid representing the probabilities of all the different combinations of values its parents could take together: (true, true), (true, false), (false, true), (false, false). The surface area of each rectangle represents the probability of that particular combination. In the example seen in Fig. 3, variable A has a 25% probability, so the vertical line has been drawn at 1/4th of the square's horizontal length. Similarly, variable B has a 75% probability, so the horizontal line has been drawn at 3/4th of the square's vertical length. The result is four rectangles, each with two parallel sides whose length corresponds to the probability of variable A being either true or false, and two parallel sides whose length corresponds to the probability of variable B being either true or false.

In these nodes, combinations that would make the node evaluate to true (such as TT, TF, and FT for an OR node) are colored green, and combinations that would make the node evaluate to false (such as FF for an OR node) are colored red. Thus, the intent is that an user can always see at a glance how probable a node is and what type of a node it is, just by looking at the ratio of green to red surface area.

When the probability of some parent node falls to 0, the rectangles representing a combination where the that variable would have been true disappear, as the length of one of their sides has been reduced to 0 (Fig. 4). Further reducing the options to just one possible eliminates all but one rectangle (now a square filling the whole node).

Sometimes rectangles in a node are colored white, to indicate that they are impossible. In Fig. 5, the user has directly observed the OR node E, and found it to be true. Because it is known to be true, it cannot have two false parents, so the FF square has been colored white.

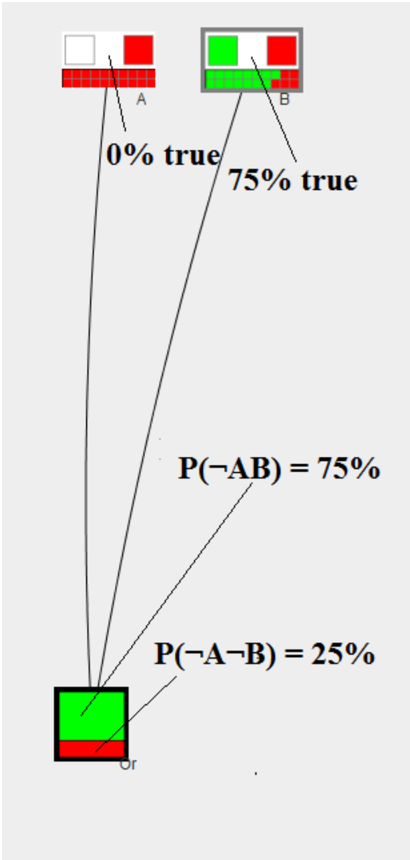



Figure 4: *An OR node with two parents A and B, with 0% and 75% probabilities respectively. Because A cannot be true, rectangles corresponding to a combination where it is true are not shown.*

### 5.1.3 Bayes nodes

A Bayes node represents a stochastic variable that has some probability of being true based on whether its parent node is true (Fig. 6). These probabilities are visualized as two grey-black bars, where the left bar represents the probability of the node being true if the parent node is true, and the right bar represents the probability of the node being true if the parent node is false. Below the grey-black bars, there are a number of green-red bars. For each full 20% chance that the parent node has of being true, a full bar will

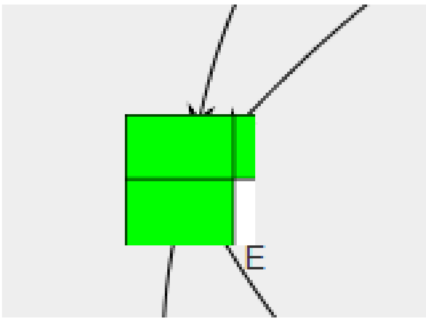


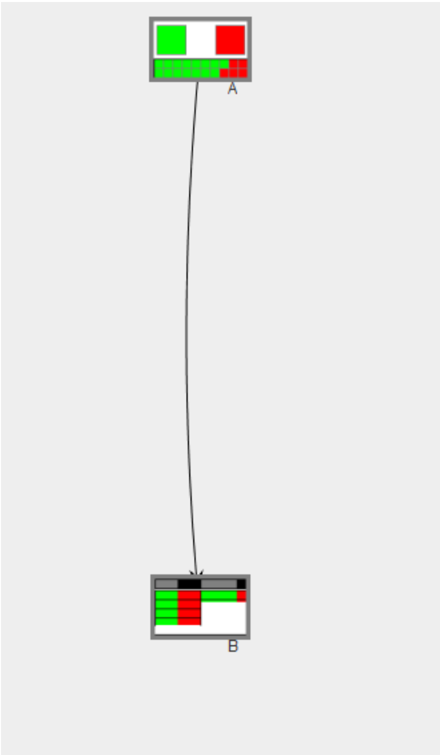

Figure 5: *An OR node that is known to be true, and thus only has green rectangles, plus one white to indicate an impossible combination of parent values.*

appear on the left side, and for each full 20% chance that the parent node has of being false, a full bar will appear on the right side. Partial bars are drawn for increments of less than 20% probability: for example, if the parent node had a 30% of being true, one and a half bars would be drawn. The balance of green and red in the bars is the same as the balance of grey and black in the bars above them.

The balance of red and green on the left side is thus a graphical representation of $P(BayesNode|Parent) \cdot P(Parent)$, and the balance of red and green on the right side a graphical representation of $P(BayesNode|\neg Parent) \cdot P(\neg Parent)$. The overall probability of the node can be read from the overall ratio of green to red.

## 5.2 Gameplay

At the start of the game, the player may choose to play or skip a brief tutorial which explains some of the basic notation of the nodes, as described in Sec. 5.1. After the tutorial, or right away if the player chose to skip the tutorial, they are presented with the message in Fig. 7. The message explains to the player the basic goal of the game: needing to raise their "psychology skill"



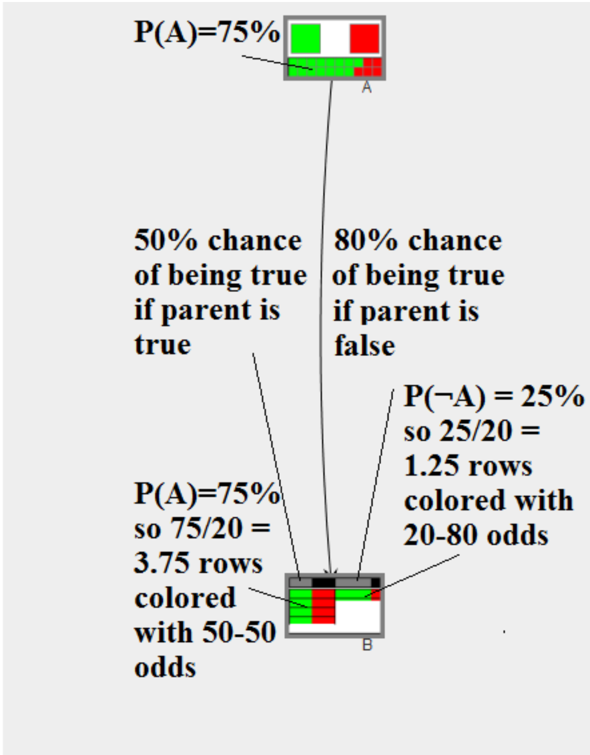



Figure 6: *A Bayes node that has a 50% probability of being true if its parent node is true, and an 80% of being true if its parent node is false. The grey bars are fixed and always represent these probabilities, while the number of colored bars depending on how likely the parent node is true or false. Each fully drawn colored bar corresponds to a 20% probability mass..*

to a specific level by each in-game day. Failure to do so will result in losing the game.

As implied by the introductory message, the game is divided into eight "days". Each day is further divided into three hours. For each hour, the player may choose to either "take an introductory lecture" or an "intermediate one" (Fig. 8). These correspond to various randomly generated Bayes nets of differing complexity.

The in-game explanation of the networks is that the lectures are given by a

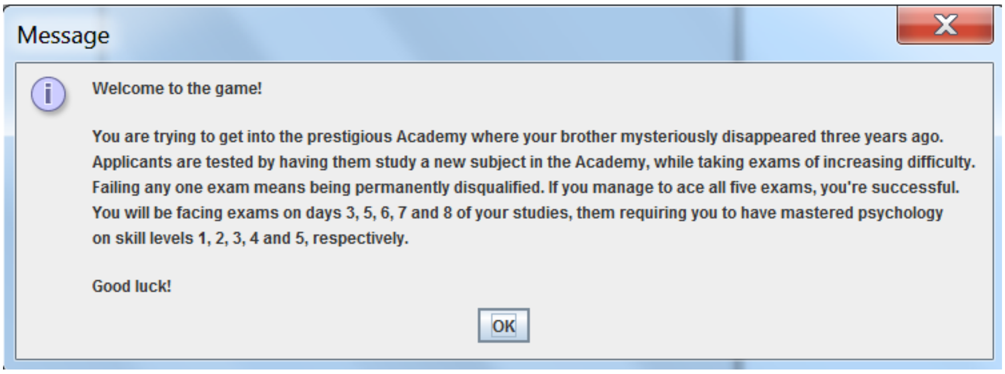



Figure 7: *An introductory message, explaining the goal of the game.*

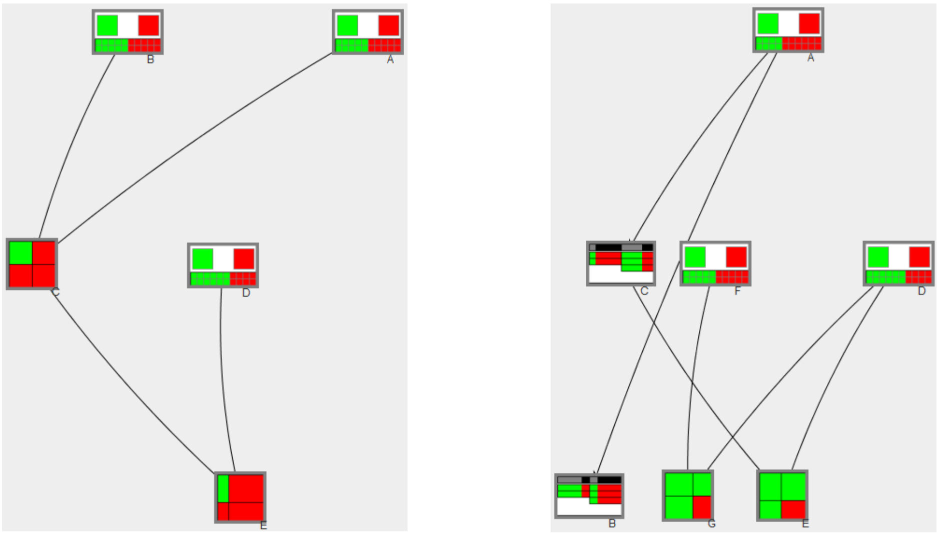


Figure 8: *Left: a sample introductory lecture. Right: a sample intermediate lecture.*

famous professor who is not very good at explaining things to students in a comprehensible way. For example, the network on the left in Fig. 8 is said to be the player character's best interpretation of the following lecture:

> 'The topic of today's lecture is observer bias. Now, it could maybe be that A. Now, it could maybe be that B. It's all about attachment, you see. Now, it logically follows that if A and B, then C. Remember what I said about observer bias? Obviously,

> it's likely that D. We figured this out because of my good friend, Dr. Prof. Weltschmerz. Now, it logically follows that if C and D, then E. (someone desperately tries to wave their hand) Straightforward, no?'
>
> 'Anyway, I hope that was enlightening. If there's anything you didn't understand, I'll leave it as a homework exercise to work it out. Thinking keeps your neurotransmitter juices flowing! Heh heh.'

Upon seeing the network, the player may either click on various nodes to observe them, or decide that they are done. Observing a node costs a point of energy. At the start of each day, the player's reserve of energy points resets to four. If the player runs out of energy, they can no longer observe any nodes.

When the player decides that they are done, they will be scored as follows. For each node in the network, the player character will assume that it will be true if it has more than a 50% probability, and false otherwise. The true values of the nodes are then revealed. If more than half of the assumptions were correct, their character will get "experience points". Accumulating sufficiently many experience points increases the player's skill level by one.

If the player manages to increase their skill level fast enough to not be rejected from the academy for failing one of the in-game "exams" (a simple check of whether or not the player character has reached the required), they will beat the game.

### 5.3 Rationale for the gameplay

The design was strongly driven by the principle of meaningful play (Sec. 3.1). In particular, the intent was that a mastery of the subject matter should be a tool towards making better decisions in the game, driving the player to want to learn the subject matter better in order to master the game. The player should be able to *discern* the consequences of their actions, and the actions should be *integrated* into the larger context of the game, affecting the play experience at later points.

The in-game lectures require the player to minimize the uncertainty of the graph while observing the smallest possible number of nodes. This requires an ability to read the network, and to identify the nodes that are the most valuable to observe. The player is never explicitly quizzed on the material, such as by asking, "which other nodes would you need to observe in order to determine the truth value of (a specific node)" or "would observing (a specific node) provide information about any other node". Yet in order to play well, the player needs to develop an implicit understanding that would let them answer such questions.

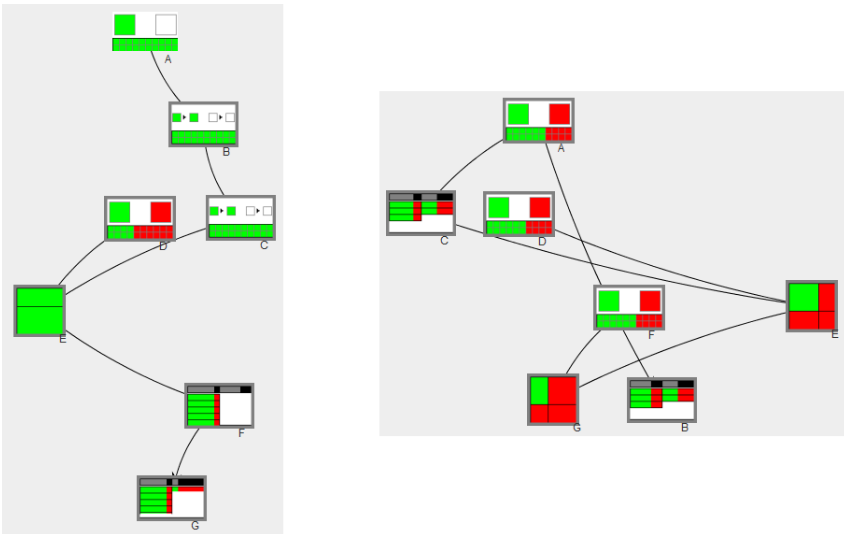


Figure 9: *Left: an easy lecture in which a single observation has paid off in many experience points. Right: a hard lecture, in which observing nodes seems unlikely to produce much of a return.*

Given the limited amount of energy per day, it may sometimes be better to inspect a very small number of nodes. For example, Fig. 9 shows two lectures from the game. In the lecture on the left, the player has observed the prior node "A", and thus found out the values of the nodes "B", "C" and "E" for sure, and the value of node "F" with a high confidence. The main uncertainties are nodes "D" and "G". In this network, the player acquired a large amount of information with just one observation, and may wish to save the rest of their energy for other networks. On the other hand, there is a bonus for getting the values of all the nodes correct, so the player may also choose to observe at least one more node.

In contrast, the lecture on the right is a much harder one. It has three prior nodes whose probabilities are all close to 50%, meaning that there is a high chance for them to be guessed wrong in the final scoring. Furthermore, there are two Bayes nodes which both have relatively close to a 50% probability of being true regardless of the truth value of their parent, so they too would need to be directly observed in order to have a high certainty of getting them right. Given that the player only gets experience points for getting at least half the network correct, it may be reasonable for the player to decide to cut their losses and not observe any nodes on this lecture, saving their energy to another lecture where spending it will hopefully produce more of a return. Thus, the need for the player to consider their remaining energy points, as well as the likely experience point gain, makes the consequences of their actions highly integrated to the game's context.

The school theme is a remainder from the original, much more ambitious design for the game. This design, similar to various "life simulator" games like Princess Maker 2 (Gainax, 1991), Magical Diary (Hanako Games, 2011), and Long Live the Queen (Hanako Games, 2012), would have required the player to split their time and energy between many more options, such as choosing between spending time with friends, going to parties, studying, and investigating the mystery of the player character's brother's disappearance. However, nearly all of these elements ended up being cut due to time constraints.

An attempt was also made to implement the principle of cognitive apprenticeship 3.2, mostly by the tutorial guiding the player through a series of tasks that started from very simple ones, such as "click on this node to observe it" and "click on this node to observe it and to find out the truth value of the IS node it is connected to". However, the tutorial was very short. Furthermore, the tutorial did not properly follow the principles of cognitive apprenticeship, as the player did not get to do much thinking of their own, but was hand-held throughout the tutorial. An alternative tutorial that would have been better in line with cognitive apprenticeship was worked on, but ultimately cut due to time constraints. The failure to properly include these principles is a likely reason why many players ended up finding the game frustrating and difficult to understand (see Sec. 6.2).

# 6 Study

Any novel educational intervention should be evaluated in order to establish whether or not it has been successful in teaching the students. Ennemoser (2010) establishes three practical questions for measuring effectiveness: 1) Is the magnitude of effects important in practical terms? 2) Are the effects durable? 3) Are there transfer effects as intended?

For Bayes Academy, the intent was that the game would also be fun. Measures for the amount of fun that the players are having can include tracking the amount of time played, the amount of levels completed, as well as subjective measures such as asking the players to rate the game on a numerical scale.

For procedural learning goals, the players may be asked to solve different problems testing the various skills involved. Transfer can be tested by giving the players problems with a different surface appearance from the problems in the game. Conceptual understanding is harder to measure, but the players may be asked to verbally explain what they feel they have learned, and to give definitions to any terms introduced in the game.

Players may also be asked how much they feel they have learned, but this is problematic. Games suffer from the risk of learners conflating the amount of fun that they have had and the amount of things that they have actually learned. Gosen & Washbush (1999) found no correlation between the self-reported perceived learning and objective measures of learning in a study of business simulation players.

I decided to test the game by recruiting people to fill out a presurvey, play the game, and the fill out a postsurvey. The surveys would offer a concrete procedural test of whether the players had learned anything, and offer a freeform “what did you learn” form to test self-reported conceptual understanding. As an oversight, the surveys did not explicitly ask whether the players had enjoyed game, but they did include another freeform “anything else you want to say” field in which people could comment on their enjoyment of the game.

## 6.1 Participants and method

I posted about the game

- on my public Facebook[5] and Google+[6] profiles
- the “Edugames” Facebook group[7]

[5]https://www.facebook.com/Xuenay
[6]https://plus.google.com/+KajSotala/posts
[7]https://www.facebook.com/groups/edugames/

- Less Wrong[8], a forum for philosophy discussions which sometimes involve Bayesianism
- the private alumni mailing list for the Center for Applied Rationality, a US nonprofit seeking to apply decision and cognitive science in people's daily lives
- a number of small IRC channels

On each venue, I linked to a pretest (appendix A) hosted on Google Forms. People who completed the pretest were given a link to download the game, as well as a link to posttest (appendix B) that they were asked to fill out once they stopped playing. The game was offered for download either as a .jar file or, for Windows users, as a package containing a copy of the JRE version the game had been developed on.

After some initial responses indicated that the objective and optimal strategy for the game was difficult to understand, three hints were added to the game. The posttest response form was then modified to include an additional question asking whether the responder had played a version of the game with the hints or without them. The game had also initially been designed for a higher resolution than many users were capable of running, causing the text to be too large and some of the graphs to be only partially visible. An updated version of the game, allowing users to choose settings adapted to lower resolutions, was released 14 hours after the initial posts.

Answers were collected from the 22nd to the 30th of September.

The pre- and post-test measured people's understanding of d-separation. The d-separation questions were identical on both forms, and people were asked not to look up the right answers from anywhere else than the game. The post-test asked how far people had gotten in the game, how much time

[8] http://lesswrong.com/

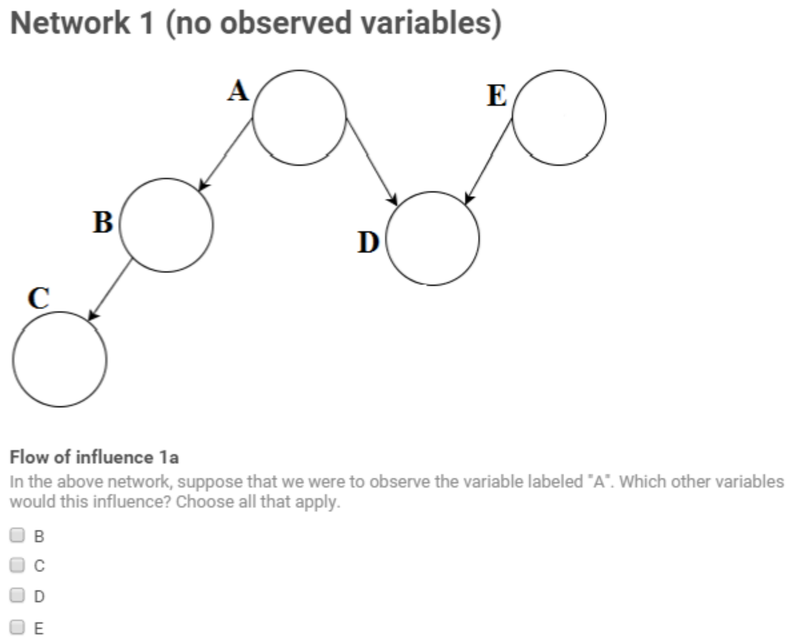



Figure 10: *A sample d-separation question from the tests. The full questionnaires are found as appendixes A and B.*

they had spent playing it, and also included two text fields labeled as "What else did you learn?" and "Anything else you want to say?".

The questions were scored by assigning one point to each checkbox that was either a) marked or b) left unmarked correctly. For example, in question 1a, one could get the full four points by marking checkboxes B, C and D, while leaving E unmarked. Because leaving checkboxes unmarked would sometimes be the right answer, even users who left all checkboxes empty would score 11 points on the tests.

To help test whether the players actually learned something conceptual that would transfer rather than only picking up on surface features, the test questions were intentionally written to require a bit of thought even for someone who had successfully mastered the game. The exact meaning of the term "influence" was never defined in the game, and the graphs in the test were drawn in a slightly different manner than in the game.

## 6.2 Results

There were 195 answers submitted to the pretest form, and 42 to the posttest form (21% retention rate). Technical problems contributed to the high number of respondents who did not go on to fill out the posttest form: in particular, many Mac users found the game unplayable due to bugs. Several users also found the interface clunky and frustrating, which may have reduced the responses further. Of the 42 people who submitted answers on the posttest, 34 had actually managed to play the game.

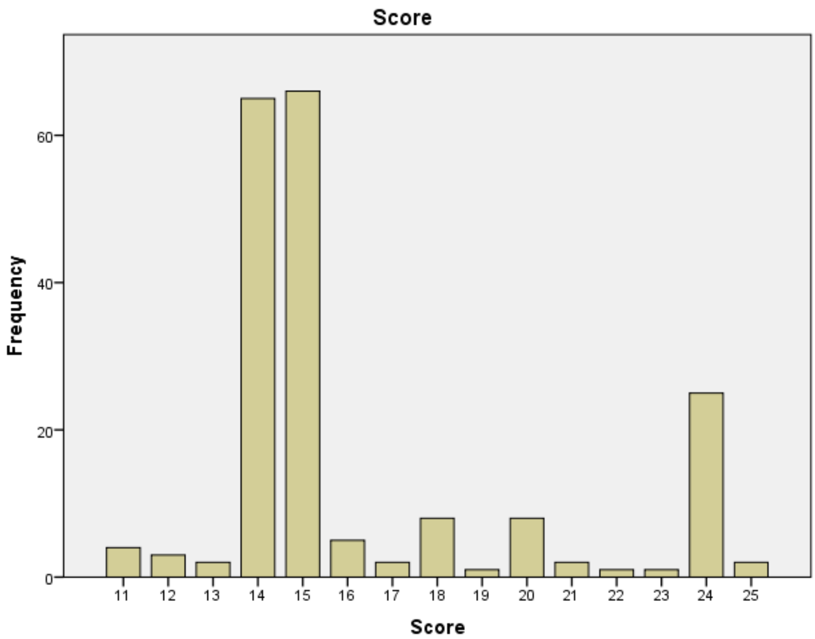



Figure 11: *Distribution of pretest d-separation scores for all users.*

A visual inspection of the pretest answers suggested the existence of two main clusters of users: people who were not previously familiar with Bayesian networks and had gotten a score of at most 15 by guessing, and people who were previously familiar with the material and scored near the maximum, 24 or 25 points.

Given the high attrition, there is the possibility that users who went on to fill out the posttest form might be a nonrepresentative sample, such as by being composed mainly of people who were already skilled with the material and found the game easier to master. However, the median pretest score was 15 for both the posttest respondents as well as the sample of all respondents, suggesting them to be similar with regard to previous experience[9].

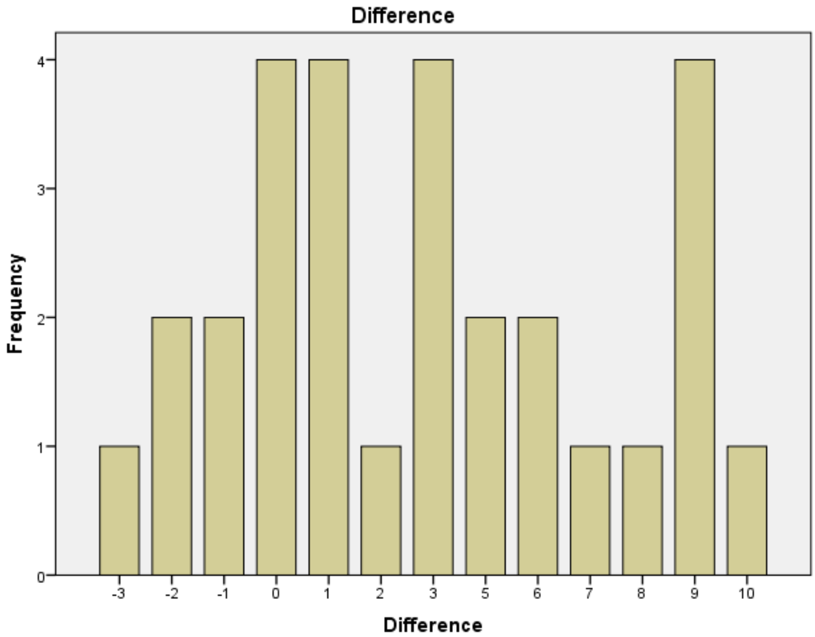



Figure 12: *Differences between posttest and pretest scores for various players. Positive difference scores indicate improvement. Players who either obtained a pretest score of 24 or above, or could not play the game for technical reasons, have been excluded.*

[9]Only median scores are reported rather than mean ones, as the tests had overlapping questions and individual test items had varying difficulties. Because of these reasons, one cannot establish a relative degree of difference between scores, making them ordinal values for which the mean is not a valid measure.

For analyzing the 34 people who had managed to play the game, I first excluded 5 users who were already familiar with the material (had scored 24 or 25 points in the pretest survey). The remaining 29 people showed an increase in test scores that ranged from -3 to 10 points, with a median improvement of 3 points (Fig. 12). A Wilcoxon signed-rank test[10] for related samples showed the difference between pre- and posttest scores to be statistically significant (p = .001), suggesting that the game had been successful in conveying an improved understanding of d-separation. Predictably, users who had started out with the lowest scores in the pretest had the most variance in their amount of improvement (Fig. 13).

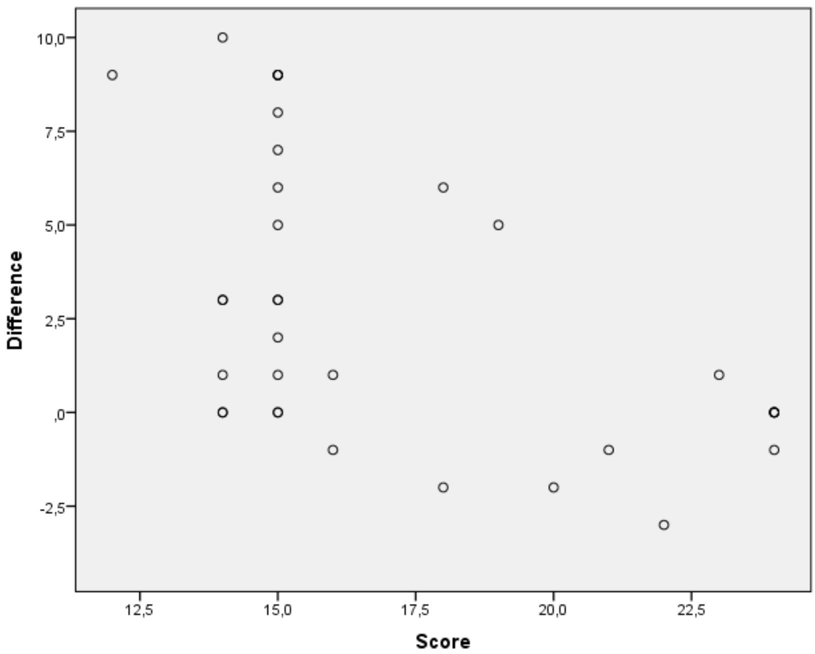



Figure 13: *Differences between posttest and pretest scores for various players, relative to their pretest scores. Positive difference scores indicate imporvement. Players who could not play the game for technical reasons have been excluded.*

[10]Used rather than a Student's t-test as the scores were clearly not normally distributed.

People who filled out the posttest form had spent between 5 and 170 minutes playing the game, with a median of 30 and a mean of 39 minutes. There was a correlation ($r_s$ = .397, p = .033, two-tailed) between the amount time that a user had spent playing the game, and the amount that their score had increased

Various responses to the open-ended "what did you learn" questions included several mentions of players learning that observing a node also affects the probabilities of its ancestors[11]:

> - I feel like I learned that we can extract information not only from the expected direction of influence, but also on the reverse. *Pretest 14, posttest 18.*

> - That the flow of influence can travel up the graph and down again. How observing a combination node changes the way the influence flows. *Pretest 15, posttest 24.*

> - Reminded myself how information can propagate "backwards". That wet street doesn't cause rain, but it does mean that it is more likely that it rained. That observing D can sometimes influence all the other values (if D is an AND and true, or if D is an OR and false (well, one of those plus B and C having good correlations with their predecessors)) *Pretest 15, posttest 24.*

Others included mentions of specific game strategies:

> - click the most uncertain boxes that affects the greatest number of other boxes. *Pretest 24, posttest 24.*

And some reported having learned nothing. In particular, one player wrote an extended, frustrated piece of feedback about it being difficult to apply the lessons from the game to the survey questions. Despite the player's

[11] More responses to this question can be found in Appendix C.1.

unhappiness with the game, I found this a particularly valuable piece of feedback. It indicated that making the transition from the in-game notation to the survey's more general diagrams required nontrivial effort, suggesting that the players who did improve their survey scores were exhibiting some measure of transfer of learning.

> - nothing. The game was completely unclear and I have no idea what the definitions are. What does it mean to be "influenced"? Does that refer to one variable affecting the value of another, or to my knowledge of one variable affecting my knowledge of another? What does it mean to be "observed"? If a variable is not observed, what does it mean to observe it? Is the game connected to these networks? It doesn't seem like it. In the game, some lines were directional (had arrows) and other were not. Why? Here all the lines have arrows. What do the various logical relations in the game have to do with these diagrams? I can't answer the questions simply because I have no idea what you're asking. It seems unfathomable to me that the game would clarify this. Someone needs to tell me what the words mean. I don't see how I could simply infer that. *Pretest 15, posttest 16.*

Ideally, progress in an educational game would directly mirror the player's skill level, so that getting to a higher level in the game would be a sign of the player having mastered the material. Conversly, a player who was familiar with the material should be able to easily advance in the game. I examined whether or not this was the case by comparing the highest level that the players reported having reached, with their score in the posttest. For this analysis I also included the players who had already been familiar with the material before the game.

The data provided partial support for mastery of the material and success in

Bayes Academy being associated (Fig. 14). Players who achieved a score of 24+ in the posttest tended to reach the later exams, and this skill level had the greatest number of players managing to beat the game. However, there was a single player who managed to beat the game despite only achieving a score of 15 in the posttest. In general, the correlation between posttest score and the level achieved fell just short of statistical significance ($r_s$ = .282, p = .053, one-tailed).

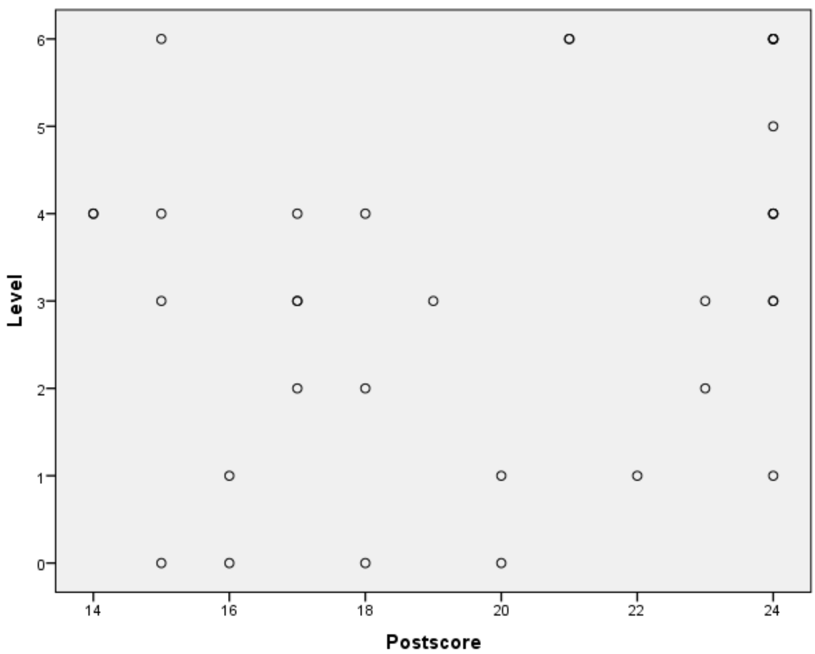



Figure 14: *In-game success for players with various scores on the posttest. Answers to the question "What was the furthest in-game day that you reached?" have been recoded into numeric values as follows: Didn't make it to day 3 = 0, Start of day 3 (Skill 1) = 1, Start of day 5 (Skill 2) = 2, Start of day 6 (Skill 3) = 3, Start of day 7 (Skill 4) = 4, Start of day 8 (Skill 5) = 5, Passed all the exams = 6*

The lack of a strong association should not be too surprising since the "exams" were not specifically designed to test the skills of the players, but rather to just make the game challenging. Although they were designed to be beatable by a strong player with some luck, reaching the later exams may have required too much luck and thus reduced the success rate of the more skilled players. Technical and interface issues likely also reduced the correlation between skill and performance. One player, who obtained a score of 24 on both the pre- and posttest, and who reached the exam on day 7 reported that:

> Game crashed a couple of times..I was on my way to finish it, but didn't bother clicking through it for the third time :(

While the game was successful in teaching players, the interface was generally considered frustrating. Several players also found the game to be confusing and not very fun, though some indicated that they had enjoyed it. Much of this was probably due to technical issues that caused the game's interface to appear differently on different platforms than the one the game had been designed on, and had effects such as often making the text very hard to read. It seems reasonable to assume that players who found the game enjoyable were overrepresented in the feedback, with players who found the game unfun or frustrating not being as likely to fill the posttest form.

Some selections from the freeform feedback[12]:

> - Thank you! Was nice to think abstract stuff without much help.
>
> - I had problems understanding the goal of the game so I never was able to get any further than my first exam. That is despite starting the game several times and carefully reading the instructions. I wasn't sure whether there was anything else to do but click on the rectangles to see the true value of the variables.
>
> - While I appreciate the effort that someone is attempting to

[12] More responses to this question can be found in Appendix C.2.

> make this, it is rather user-unfriendly, which somewhat defeats the point. ... All in all, I had a fairly frustrating experience with the UI, and stumbled my way through a few levels before making it unwinnable. Despite all that, through the bugs, I can see the framework of a game I've dreamed of for years. I hope my list above is not too discouraging; I'd really like to see this game finished!

# 7 Conclusions

Despite being generally considered unintuitive and confusing, Bayes Academy produced some promising results in teaching its players basic concepts about Bayesian networks. I will now consider the extent to which the results answer Ennemoser (2010)'s three practical questions for measuring the effectiveness of an educational game.

**1) Is the magnitude of effects important in practical terms?**

While the results were clearly statistically significant, Ennemoser emphasizes that statistical significance does not equal practical significance. In this respect, the results were clearly varied: a number of players did not improve at all, while a few actually performed worse on the posttest than on the pretest. The median improvement was only 3 points. On the hand, there were players who demonstrated a considerable increase in skill, and the players who had performed the worst initially had the largest chance of making a considerable improvement.

Because the pretest and posttest questions were intentionally written to be somewhat vague so as to test transfer of learning and conceptual understanding, it seems plausible that these results might underestimate the amount of learning that happened.

**2) Are the effects durable?**

Ennemoser recommends testing this by having a follow-up study later on. Unfortuntately, the scope of this study did not allow for this, so the durability of the effects remains undetermined.

**3) Are there transfer effects as intended?**

While the participants seem unlikely to have learned much that would have transferred to the context of daily life, the fact that many showed an improvement on a test which was dissimilar to the game suggests that some transfer of learning did indeed take place. The assumption of this having been a nontrivial task seems supported by the frustrated feedback by one of the players, who found it too difficult to translate the game's notation into the notation and vocabulary used in the posttest. Thus, it seems like some amount of transfer of learning happened to at least some of the players, hopefully leading to a more general understanding.

While being moderately educational, for most players the game clearly failed in its goal of being fun. While the feedback did include positive reactions to the game, many people also reported an overall negative reaction, and it seems likely that people who reacted positively were overrepresented among the respondents. The lack of fun was likely to a number of reasons, including an inadequately explained interface, technical issues, failure to properly follow the principles of cognitive apprenticeship, too high of an initial difficulty, poor explanations of what was going on, and the consequences of the player's actions not being discernable enough.

Despite these shortcomings, the project has demonstrated that educational games have promise in teaching technical topics such as Bayesian networks. Many of the flaws could be fixed with more work, and I am already considering a possible follow-up for the project.

# References

Bennie, K. (1998). *The 'slippery' concept of probability: reflections on possible teaching approaches.* Paper presented at 4th Annual Congress of the Association for Mathematics Education in South Africa (AMESA), Pietersburg.. Retrieved 2013-09-29, from `http://academic.sun.ac.za/mathed/malati/files/Probability98.pdf`

Boondao, R. (2008, May). Crime risk factors analysis. In O. Pourret, P. Naïm, & B. Marcot (Eds.), *Bayesian Networks: A Practical Guide to Applications* (1st ed.). Chichester, West Sussex, England; Hoboken, New Jersey: Wiley.

Bruckman, A. (1999). *Can educational be fun.* Paper presented at Game Developers Conference 1999, San Jose, California. Retrieved 2015-11-10, from `http://www.cc.gatech.edu/~asb/papers/conference/bruckman-gdc99.pdf`

Caillois, R. (1961). *Man, Play, and Games.* New York: Free Press.

Campos, L. M. d., Fernández-Luna, J. M., Huete, J. F., Martín, C., & Romero, A. E. (2008, May). An information retrieval system for parliamentary documents. In O. Pourret, P. Naïm, & B. Marcot (Eds.), *Bayesian Networks: A Practical Guide to Applications* (1st ed.). Chichester, West Sussex, England; Hoboken, New Jersey: Wiley.

Cengage Learning, & Muzzy Lane Software. (2012). *ALGEBURST Lite. [iOS Game].* Boston, Massachusetts: Cengage Learning.

Chater, N., Tenenbaum, J. B., & Yuille, A. (2006). Probabilistic models of cognition: Conceptual foundations. *Trends in Cognitive Sciences*, *10*(7), 287–291.

Collins, A. (1988). *Cognitive Apprenticeship and Instructional Technology* (Technical Report). University of Illinois at Urbana-Champaign.

Collins, A. (2006). Cognitive Apprenticeship. In R. Sawyer (Ed.), *The Cambridge Handbook of the Learning Sciences* (pp. 47–60). New York: Cambridge University Press.

Collins, A., Brown, J. S., & Holum, A. (1991). Cognitive apprenticeship: Making thinking visible. *American Educator*, *6*(11), 38–46.

Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011). From game design elements to gamefulness: defining gamification. In *Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments* (pp. 9–15). New York: ACM New York.

Duarte-Mermoud, M. A., Beltrán, N. H., & Vergara, S. H. (2008, May). Classification of Chilean wines. In O. Pourret, P. Naïm, & B. Marcot (Eds.), *Bayesian Networks: A Practical Guide to Applications* (1st ed.). Chichester, West Sussex, England; Hoboken, New Jersey: Wiley.

Eduskunnan tulevaisuusvaliokunta. (2013). *Uusi Oppiminen* (Tech. Rep. No. 8/2013).

Ennemoser, M. (2010). Evaluating the Potential of Serious Games: What Can We Learn from Previous Research on Media Effects and Educational Intervention? In U. Ritterfeld, M. Cody, & P. Vorderer (Eds.), *Serious Games: Mechanisms and Effects* (1st ed.). New York: Routledge.

Faria, A. J. (1998, September). Business Simulation Games: Current Usage Levels—An Update. *Simulation & Gaming*, *29*(3), 295–308. doi: 10.1177/1046878198293002

Farr, C. (2013, October). The president's 'gaming guy' tells us that educational games fascinate Obama. *Venture Beat*. Retrieved 2013-12-10, from `http://venturebeat.com/2013/10/29/the-presidents-gaming-guy-tells-us-that-games-fascinate-obama/`

Firaxis Games. (2013). *XCOM: Enemy Unknown [PC Game].* Novato, California: 2K Games.

Gainax. (1991). *Princess Maker 2 [PC Game].* Mitaka, Tokyo: Gainax.

Garbolino, P., & Taroni, F. (2002). Evaluation of scientific evidence using Bayesian networks. *Forensic Science International*, *125*(2), 149–155.

Gee, J. P. (2007). *What Video Games Have to Teach Us About Learning and Literacy.* New York: Palgrave Macmillan.

Gigerenzer, G., & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review*, *102*(4), 684.

Gosen, J., & Washbush, J. (1999). Perceptions of learning in TE simulations. *Developments in Business Simulation & Experiential Exercises*, *26*, 170–175.

Graesser, A., Chipman, P., Leeming, F., & Biedenbach, S. (2010, April). Deep Learning and Emotion in Serious Games. In U. Ritterfeld, M. Cody, & P. Vorderer (Eds.), *Serious Games: Mechanisms and Effects* (1st ed.). New York: Routledge.

Greco, M., Baldissin, N., & Nonino, F. (2013, October). An Exploratory Taxonomy of Business Games. *Simulation & Gaming*, *44*(5), 645–682. doi: 10.1177/1046878113501464

Hahn, U., & Oaksford, M. (2007). The rationality of informal argumentation: A Bayesian approach to reasoning fallacies. *Psychological Review*, *114*(3), 704–732. doi: 10.1037/0033-295X.114.3.704

Hanako Games. (2011, June). *Magical Diary [PC Game].* Sussex, UK.

Hanako Games. (2012, June). *Long Live the Queen [PC Game].* Sussex, UK.

Heeter, C., Lee, Y.-H., Magerko, B., & Medler, B. (2011). Impacts of forced serious game play on vulnerable subgroups. *International Journal of Gaming and Computer-Mediated Simulations (IJGCMS)*, *3*(3), 34–53.

Hájek, A. (2012). Interpretations of Probability. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2012 ed.). Retrieved 2013-12-10, from `http://plato.stanford.edu/entries/probability-interpret/`

Hoffrage, U., Gigerenzer, G., Krauss, S., & Martignon, L. (2002). Representation facilitates reasoning: What natural frequencies are and what they are not. *Cognition*, *84*(3), 343–352.

Howson, C., & Urbach, P. (1993). *Scientific Reasoning: the Bayesian Approach.* Chicago: Open Court.

Huizinga, J. (1955). *Homo Ludens; a Study of the Play-Element in Culture.* Boston, Massachusetts: Beacon Press.

Itti, L., & Baldi, P. F. (2005). Bayesian surprise attracts human attention. In Y. Weiss, B. Schölkopf, & J. C. Platt (Eds.), *Advances in Neural Information Processing Systems 18* (pp. 547–554).

Jones, M., & Love, B. C. (2011, August). Bayesian Fundamentalism or Enlightenment? On the explanatory status and theoretical contributions of Bayesian models of cognition. *Behavioral and Brain Sciences*, *34*(04), 169–188. doi: 10.1017/S0140525X10003134

Koller, D., & Friedman, N. (2009). *Probabilistic Graphical Models: Principles and Techniques* (1st ed.). Cambridge, Massachusetts: The MIT Press.

Kording, K. (2007, October). Decision Theory: What "Should" the Nervous System Do? *Science*, *318*(5850), 606–610. doi: 10.1126/science.1142998

Koster, R. (2010). *Theory of Fun for Game Design* (1st ed.). Scottsdale, Arizona: Paraglyph Press.

Lardinois, F. (2013, June). Duolingo Comes To The iPad, Now Has More Than 5m Active Users. *TechCrunch*. Retrieved 2013-12-10, from `http://techcrunch.com/2013/07/11/duolingo-comes-to-the-ipad-now-has-more-than-5m-active-users/`

Lave, J. (1988). *Cognition in Practice: Mind, Mathematics and Culture in Everyday Life*. Boston, Massachusetts: Cambridge University Press.

Macaulay, C. (2001). Transfer of learning. In C. Macaulay & V. E. Cree (Eds.), *Transfer of Learning in Professional and Vocational Education: Handbook for Social Work Trainers* (pp. 1–26). London, UK: Routledge.

Marczewski, A. (2013, February). What's the Difference Between Gamification and Serious Games? *Andrzejs Blog [Blog]*. Retrieved 2014-01-12, from `http://marczewski.me.uk/2013/02/25/gamification-and-serious-games/`

Micallef, L., Dragicevic, P., & Fekete, J. (2012). Assessing the effect of visualizations on bayesian reasoning through crowdsourcing. *Visualization and Computer Graphics, IEEE Transactions on*, *18*(12), 2536–2545.

Michael, D. R., & Chen, S. L. (2005). *Serious games: Games That Educate, Train, and Inform*. Boston, Massachusetts: Thomson.

MTV3. (2013, September). Pelialan huiput kutsutaan ideoimaan matematiikan opetusta. *mtv3.fi*. Retrieved 2013-12-10, from `http://www.mtv.fi/uutiset/kotimaa/artikkeli/pelialan-huiput-kutsutaan-ideoimaan-matematiikan-opetusta/2334206`

Nicholson, A. E., Twardy, C. R., Korb, K. B., & Hope, L. R. (2008, May). Decision support for clinical cardiovascular risk assessment. In O. Pourret, P. Naïm, & B. Marcot (Eds.), *Bayesian Networks: A Practical Guide to Applications* (1st ed.). Chichester, West Sussex, England; Hoboken, New Jersey: Wiley.

Onisko, A. (2008, May). Medical diagnosis. In O. Pourret, P. Naïm, & B. Marcot (Eds.), *Bayesian Networks: A Practical Guide to Applications* (1st ed.). Chichester, West Sussex, England; Hoboken, New Jersey: Wiley.

Palomäki, J. (2013). *New Perspectives on Emotional Processes and Decision Making in the Game of Poker : With Special Emphasis on the Tilting Phenomenon* (PhD Thesis). University of Helsinki, Helsinki, Finland.

Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Francisco: Morgan Kaufmann.

Porwal, A., & Carranza, E. (2008, May). Classifiers for modeling of mineral potential. In O. Pourret, P. Naïm, & B. Marcot (Eds.), *Bayesian Networks: A Practical Guide to Applications* (1st ed.). Chichester, West Sussex, England; Hoboken, New Jersey: Wiley.

Rittle-Johnson, B., Siegler, R. S., & Alibali, M. W. (2001). Developing conceptual understanding and procedural skill in mathematics: An iterative process. *Journal of Educational Psychology*, *93*(2), 346.

Salen, K., & Zimmerman, E. (2004). *Rules of Play: Game Design Fundamentals*. Cambridge, Massachusetts: The MIT Press.

Schell, J. (2008). *The Art of Game Design: A Book of Lenses*. Burlington, Massachusetts: Taylor & Francis US.

Sebastiani, P., & Perls, T. T. (2008, May). Complex genetic models. In O. Pourret, P. Naïm, & B. Marcot (Eds.), *Bayesian Networks: A*

*Practical Guide to Applications* (1st ed.). Chichester, West Sussex, England; Hoboken, New Jersey: Wiley.

Sedlmeier, P. (1997). BasicBayes: A tutor system for simple Bayesian inference. *Behavior Research Methods, Instruments, & Computers*, *29*(3), 328–336.

Sedlmeier, P. (2000). How to improve statistical thinking: Choose the task representation wisely and learn by doing. *Instructional Science*, *28*(3), 227–262.

Sedlmeier, P., & Gigerenzer, G. (2001). Teaching Bayesian reasoning in less than two hours. *Journal of Experimental Psychology: General*, *130*(3), 380.

Shaffer, D. W. (2006). *How Computer Games Help Children Learn.* New York: Palgrave Macmillan.

Sherry, J. L., & Dibble, J. L. (2010, April). The Impact of Serious Games on Childhood Development. In U. Ritterfeld, M. Cody, & P. Vorderer (Eds.), *Serious Games: Mechanisms and Effects* (1st ed.). New York: Routledge.

Squire, K. (2004). *Replaying History: Learning World History through playing Civilization III* (Doctoral dissertation). Indiana University.

Squire, K. (2006). From content to context: Videogames as designed experience. *Educational Researcher*, *35*(8), 19–29.

Sterrett, J. (2012, November). Interview : James Sterrett, Professional Wargamer. *Rock, Paper, Shotgun.* Retrieved 2013-12-06, from `http://www.rockpapershotgun.com/2012/11/16/interview-james-sterrett-professional-wargamer/`

Stickgold, R. (2000, October). Replaying the Game: Hypnagogic Images in Normals and Amnesics. *Science*, *290*(5490), 350–353. doi: 10.1126/science.290.5490.350

Sucar, L. E., & Noguez, J. (2008, May). Student modeling. In O. Pourret, P. Naïm, & B. Marcot (Eds.), *Bayesian Networks: A Practical Guide to Applications* (1st ed.). Chichester, West Sussex, England; Hoboken, New Jersey: Wiley.

Talbott, W. (2013). Bayesian Epistemology. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2013 ed.). Retrieved 2013-12-10, from `http://plato.stanford.edu/entries/epistemology-bayesian/`

TVTropes. (2014, October). *Main/The Tetris Effect - Television Tropes & Idioms.* Retrieved 2014-10-22, from `http://tvtropes.org/pmwiki/pmwiki.php/{Main}/{TheTetrisEffect}`

U.S. Army, & Secret Level. (2002). *America's Army. [PC Game].* Washington D.C.: U.S. Army.

Uusitalo, L. (2007, May). Advantages and challenges of Bayesian networks in environmental modelling. *Ecological Modelling*, *203*(3-4), 312–318. doi: 10.1016/j.ecolmodel.2006.11.033

Washington State. (2013). Washington State Algebra Challenge. Retrieved 2013-12-10, from `http://wa.algebrachallenge.org/`

WeWantToKnow. (2012). *DragonBox [Mobile game].* Oslo, Norway: WeWantToKnow.

Whitton, N. (2007). *An Investigation Into the Potential of Collaborative Computer Game-Based Learning in Higher Education* (Unpublished doctoral dissertation). Napier University.

Whitton, N. (2010). *Learning With Digital Games: a Practical Guide to Engaging Students in Higher Education.* New York: Routledge.

Zahner, D., & Corter, J. E. (2010, March). The Process of Probability Problem Solving: Use of External Visual Representations. *Mathematical Thinking and Learning*, *12*(2), 177–204. doi: 10.1080/10986061003654240

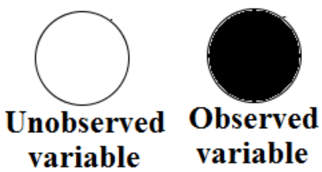
# A Appendix - Pretest form

Hey! Thanks for your interest in helping me gather data for my game. Before you play it, I'd like to ask you to answer a few questions, so that I can find out how much you've actually learned. Once you've submitted the form, you will get a link to the actual game. [The game requires Java.] This survey tests your understanding of d-separation in Bayesian networks. If you do not understand the questions, do not worry. Just make your best guess, or leave the answer blank if you genuinely have no idea.

In the pictures below, observed variables will be in black, and unobserved variables will be white. (Again, do not worry if you do not know what this means.)

NOTE: please do not look up the correct answers to these questions from anywhere else than playing the game, at least until you've filled out the post-test.

* Required




Your name *

Please enter a name from which we can recognize you later on, when you fill out the game post-test. This doesn't have to be your real name, any string of symbols which you believe will be unique among the participants and which you'll remember will do.

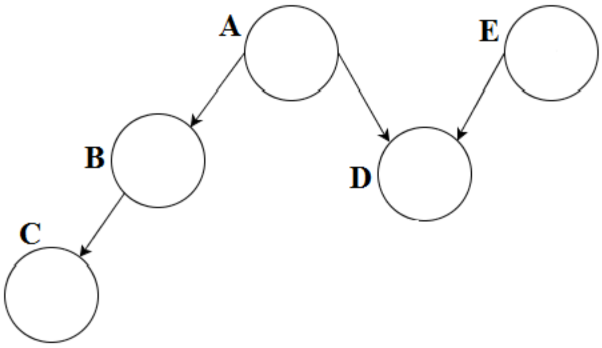
Previous game exposure

Have you read any of Kaj's earlier blog posts about the "Bayes Academy" game?

☐ Ones on kajsotala.fi

☐ Ones on lesswrong.com

**Network 1 (no observed variables)**




Flow of influence 1a

In the above network, suppose that we were to observe the variable labeled "A". Which other variables would this influence? Choose all that apply.

☐ B

☐ C

☐ D

☐ E

Flow of influence 1b

In the above network, suppose that we were to observe the variable labeled "B". Which other variables would this influence? Choose all that apply.

☐ A

☐ C

☐ D

☐ E

Flow of influence 1c

In the above network, suppose that we were to observe the variable labeled

"C". Which other variables would this influence? Choose all that apply.

☐ A

☐ B

☐ D

☐ E

Flow of influence 1d

In the above network, suppose that we were to observe the variable labeled "D". Which other variables would this influence? Choose all that apply.

☐ A

☐ B

☐ C

☐ E

Flow of influence 1e

In the above network, suppose that we were to observe the variable labeled "E". Which other variables would this influence? Choose all that apply.

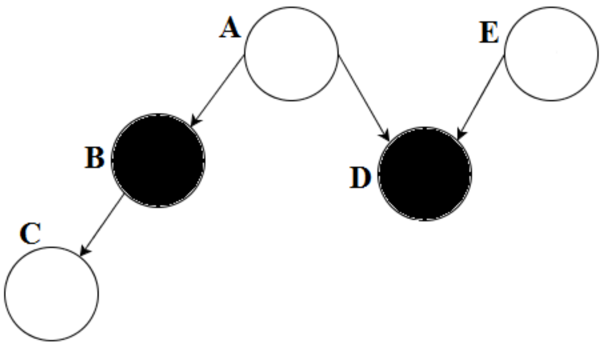
☐ A

☐ B

☐ C

☐ D

**Network 2 (B, D observed)**




Flow of influence 2a

In the above network, suppose that we were to observe the variable labeled "A". Which other variables would this influence? Choose all that apply.

☐ C

☐ E

Flow of influence 2c

In the above network, suppose that we were to observe the variable labeled "C". Which other variables would this influence? Choose all that apply.

☐ A

☐ E

Flow of influence 2e

In the above network, suppose that we were to observe the variable labeled "E". Which other variables would this influence? Choose all that apply.

☐ A

☐ C

# B Appendix - Posttest form

Hey! Thanks for playing my game. Now that you've done so, please answer a few quick questions.
Below are some of the same questions that you answered in the pretest. Hopefully you might now have a bit better of an understanding of them. (If not, again don't worry - just make your best guess, or leave the answer blank if you genuinely have no idea.)

In the pictures below, observed variables will be in black, and unobserved variables will be white.

* Required

Your name *
Here, type in the name that you entered in the pretest form.

Approximately how much real-life time did you spend on the game? *
(Just ignore the "seconds" field)

*(drop-down form with the fields "hours" : "minutes" : "seconds")*

What was the furthest in-game day that you reached? *

O Didn't make it to day 3
O Start of day 3 (Skill 1)
O Start of day 5 (Skill 2)
O Start of day 6 (Skill 3)
O Start of day 7 (Skill 4)
O Start of day 8 (Skill 5)
O Passed all the exams

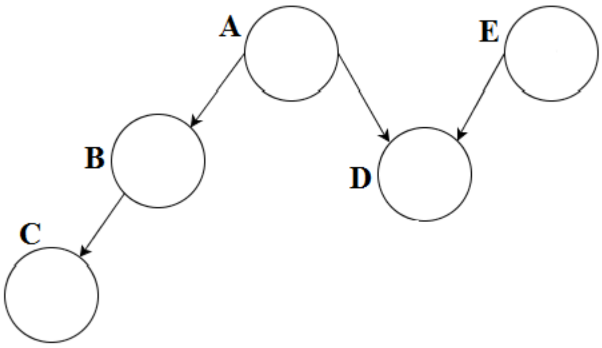
Did you cheat and look up how you should answer any of the "flow of influence” questions below, from any other source than the game you played? *

O No sir! O Sure did!

Did you play the version that did or didn’t have the three hints displayed at the beginning of the first lecture? *

O Saw the hints O No, what hints?

**Network 1 (no observed variables)**




Flow of influence 1a

In the above network, suppose that we were to observe the variable labeled “A”. Which other variables would this influence? Choose all that apply.

☐ B

☐ C

☐ D

☐ E

Flow of influence 1b

In the above network, suppose that we were to observe the variable labeled “B”. Which other variables would this influence? Choose all that apply.

☐ A

☐ C

☐ D

☐ E

Flow of influence 1c

In the above network, suppose that we were to observe the variable labeled "C". Which other variables would this influence? Choose all that apply.

☐ A

☐ B

☐ D

☐ E

Flow of influence 1d

In the above network, suppose that we were to observe the variable labeled "D". Which other variables would this influence? Choose all that apply.

☐ A

☐ B

☐ C

☐ E

Flow of influence 1e

In the above network, suppose that we were to observe the variable labeled "E". Which other variables would this influence? Choose all that apply.

☐ A

☐ B

☐ C

☐ D

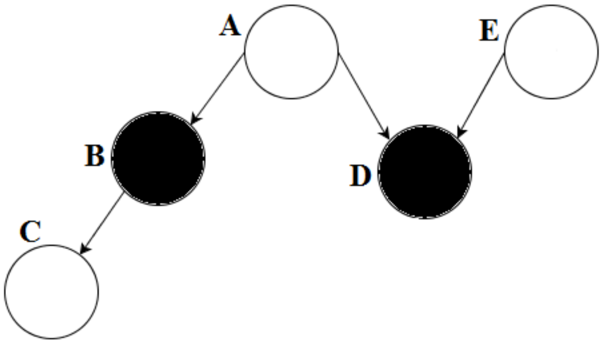
**Network 2 (B, D observed)**

Flow of influence 2a

In the above network, suppose that we were to observe the variable labeled "A". Which other variables would this influence? Choose all that apply.

☐ C

☐ E




Flow of influence 2c

In the above network, suppose that we were to observe the variable labeled "C". Which other variables would this influence? Choose all that apply.

☐ A

☐ E

Flow of influence 2e

In the above network, suppose that we were to observe the variable labeled "E". Which other variables would this influence? Choose all that apply.

☐ A

☐ C

What did you feel you learned?

Anything else you want to say?

# C Appendix - Extended selection of feedback responses

## C.1 "What did you learn" question

Various responses to the open-ended "what did you learn" questions included the following:

> - I feel like I learned that we can extract information not only from the expected direction of influence, but also on the reverse. *Pretest 14, posttest 18.*

> - nothing. The game was completely unclear and I have no idea what the definitions are. What does it mean to be "influenced"? Does that refer to one variable affecting the value of another, or to my knowledge of one variable affecting my knowledge of another? What does it mean to be "observed"? If a variable is not observed, what does it mean to observe it? Is the game connected to these networks? It doesn't seem like it. In the game, some lines were directional (had arrows) and other were not. Why? Here all the lines have arrows. What do the various logical relations in the game have to do with these diagrams? I can't answer the questions simply because I have no idea what you're asking. It seems unfathomable to me that the game would clarify this. Someone needs to tell me what the words mean. I don't see how I could simply infer that. *Pretest 15, posttest 16.*

> - I feel like I only learned that the observation of variables has a greater influence than expected, i.e. that predecessors are influenced as well. *Pretest 15, posttest 22.*

> - I did have some understanding of bayes nets before - but I don't

think I learned anything concrete. Interface might be the problem - only about in the middle of the game I could differentiate between IS-node and normal node, arrows were sinking into nodes. *Pretest 15, posttest 15.*

- click the most uncertain boxes that affects the greatest number of other boxes. *Pretest 24, posttest 24.*

- No hard facts, but the iteration of intellectually known things should help with making them intuitively known. *Pretest 24, posttest 24.*

- I already knew that in a case such as "If A, then B" knowledge about B would influence A, but I didn't think of it answering the question before I had played the game. *Pretest 14, posttest 24.*

- Understanding of how to read and interpret a specific jargon of diagram *Pretest 15, posttest 24.*

- Not much. *Pretest 24, posttest 23.*

- In the pre-test, I wasn't sure what "influence" meant. I interpreted the circles and arrows to mean causality, like when Eliezer draws circles for it being raining and the sidewalk being wet and the sprinkler being on, and there are arrows from the sprinkler being wet and the sidewalk being on. And I interpreted "influence" as also being a causality statement, which didn't seem right, but that's how I answered the question.

When playing the game, it was clear that the point was, given that you learn one variable, what do you learn about the others, which caused me to reinterpret influence during the post-test as "which variables would you learn something new about if you observed this variable". *Pretest 15, posttest 19; due to technical* *difficulties, could only play first five minutes of game.*

- That the flow of influence can travel up the graph and down again. How observing a combination node changes the way the influence flows. *Pretest 15, posttest 24.*

- I feel I learned that probabilities can update in reverse direction, and that sometimes you get unlucky when you play the probabilities. *Pretest 15, posttest 18.*

- Not much? I'm still unclear what exactly you mean by 'flow of influence'. For the pre-questions I answered it like the arrows were causal but not necessarily related to other arrows from the same node, which didn't match the simple nodes you used in the game. I think I am more confused rather than less. *Pretest 15, posttest 23.*

- Reminded myself how information can propagate "backwards". That wet street doesn't cause rain, but it does mean that it is more likely that it rained. That observing D can sometimes influence all the other values (if D is an AND and true, or if D is an OR and false (well, one of those plus B and C having good correlations with their predecessors)) *Pretest 15, posttest 24.*

- Not much :( More than the actual game, I was trying to figure out the notation. The game made the idea of 'value of information' a bit more salient. If information were free, I'd have had unlimited 'energy' (being allowed to look up truth of nodes). *Pretest 20, posttest 18.*

- When you reveal a node information won't traverse it any longer.

When you reveal a node, the probabilities of its ancestors and successors are updated.

When you reveal a node, you split the network (by stopping information traversal). You'd probably want to plan for this, I didn't.

Well-connected nodes can move the probabilities of a lot of nodes, but you'll still need to catch those 0.5-0.5 buggers somehow.

I wanted to map the different networks and try by experiment which nodes were good to reveal. Also wanted to know how many points you need in general to solve each network, and how many points to use on each type of network. *Pretest 22, posttest 19.*

- Kaj Sotala's game programming skills are not quite as good as his blogging skills. *Pretest 18, posttest 16.*

## C.2 "Anything else you want to say" question

- Right now, the game is not very fun, but it has the potential to be fun. I hope it achieves that potential. (Ignoring for a moment the opportunity cost of you working on this instead of on much cooler stuff.)

- Well, i'm sorry but after a few try i did not find how to "strategically" play so i just adopt a gaming strategy to clic on the square above others with 50/50 luck and it seem to return not bad result with minimum involvement for me and some luck. But i don't really understand why.

- I'm sorry, but the UI was terrible :(

- Thank you! Was nice to think abstract stuff without much help.

- I had problems understanding the goal of the game so I never was able to get any further than my first exam. That is despite starting the game several times and carefully reading the instructions. I wasn't sure whether there was anything else to do but click on the rectangles to see the true value of the variables.

- While I appreciate the effort that someone is attempting to make this, it is rather user-unfriendly, which somewhat defeats the point. ... All in all, I had a fairly frustrating experience with the UI, and stumbled my way through a few levels before making it unwinnable. Despite all that, through the bugs, I can see the framework of a game I've dreamed of for years. I hope my list above is not too discouraging; I'd really like to see this game finished!

- on mac os this game was truly not usable; even with the lowest resolution the text didn't always fit on the screen (720 with menu bar, not without!), on other occasions parts of the game covered one another or fonts were displayed in unreadable sizes. Without tutorial I did not really figure out what the "intermediate course" symbols meant. maybe exclude this from data

- Interesting concept, really rough UI. No idea how the Bayes cells worked. Not really clear on what I needed to do to win, it's a resource management game where you need to get maximum certainty with minimal energy, instead of a thing where you have a clear condition of ""everything is same color, you win the round""?

How much variance do you expect on the graph answers? I guess you'll get ""overthinking it"" flips on which way you expect the arrows to act, and whether the effect is transitive or just affects the immediate next node.

- I only played once [to me, it seemed like the fairest way, though the probabilistic nature of the task probably voids that particular heuristic]. I really enjoyed it.

- I was playing on a Mac, and I'm not sure what you meant by the info box, but assuming it meant the box where writing appears at the bottom, that did indeed disappear after a few steps of the tutorial. (However, it wasn't at the top in the main game, so maybe you meant something else.) I couldn't understand the rules of the game, probably because I couldn't understand the tutorial, so I didn't get very far. I clicked one box on the first day, then wanted to go to the other of the two courses, so I clicked done. Then it said I had a score of 0 and it was the next day, so I got frustrated and stopped playing.

If you ever get this game working for Mac, I'd love to play!